基于无人机遥感和深度学习的葡萄卷叶病感染程度诊断方法

刘易雪[1,2,3]，宋育阳[4]，崔　萍[5]，房玉林[4]，苏宝峰[1,2,3*]

（1. 西北农林科技大学 机械与电子工程学院，陕西杨凌 712100，中国；2. 农业农村部农业物联网重点实验室，陕西杨凌 712100，中国；3. 陕西省农业信息感知与智能服务重点实验室，陕西杨凌 712100，中国；4. 西北农林科技大学 葡萄酒学院，陕西杨凌 712100，中国；5. 宁夏贺兰山东麓葡萄产业园区管理委员会，宁夏银川，750002，中国）

摘　要：［**目的/意义**］葡萄卷叶病是一种严重影响葡萄产量和品质的病害。葡萄卷叶病感染程度类别之间存在严重的数据不平衡，导致无人机遥感技术难以进行精确的诊断。针对此问题，本研究提出一种结合细粒度分类和生成对抗网络（Generative Adversarial Network，GAN）的方法，用于提高无人机遥感图像中葡萄卷叶病感染程度分类的性能。［**方法**］以蛇龙珠品种卷叶病识别诊断为例，使用GANformer分别对每一类的葡萄园正射影像的分块图像进行学习，生成多样化和逼真的图像以增强数据，并以Swin Transformer tiny作为基础模型，提出改进模型CA-Swin Transformer，引入通道注意力机制（Channel Attention，CA）来增强特征表达能力，并使用ArcFace损失函数和实例归一化（Instance Normalization，IN）来改进模型的性能。［**结果和讨论**］GANformer可以生成FID score为93.20的蛇龙珠虚拟冠层图像，有效地改善数据不平衡问题。同时，相比基于卷积神经网络（Convolutional Neural Networks，CNN）的深度学习模型，基于Transformer的深度学习模型在卷叶病感染程度诊断的问题上更具优势。最佳模型Swin Transformer在增强数据集上达到83.97%的准确率，比在原始数据集上提高3.86%，且高于GoogLeNet、MobileNetV2、NasNet Mobile、ResNet18、ResNet50、CVT和T2TViT等对照模型。而本研究所提的CA-Swin Transformer在增强数据后的测试集上达到86.65%的分类精度，比在原始的测试集上使用Swin Transformer精度提高6.54%。［**结论**］本研究基于CA-Swin Transformer使用滑动窗口法制作了葡萄园蛇龙珠卷叶病严重程度分布图，为葡萄园卷叶病的防治提供了参考。同时，本研究的方法为无人机遥感监测作物病害提供了一种新的思路和技术手段。

关键词：无人机遥感；深度学习；生成对抗网络；Swin Transformer；酿酒葡萄卷叶病；数据增强；注意力机制

中图分类号：TP18;TP391.41;S436.68　　**文献标志码：**A　　**文章编号：**SA202308013

引用格式：刘易雪, 宋育阳, 崔萍, 房玉林, 苏宝峰. 基于无人机遥感和深度学习的葡萄卷叶病感染程度诊断方法[J]. 智慧农业(中英文), 2023, 5(3): 49-61. DOI：10.12133/j.smartag.SA202308013

LIU Yixue, SONG Yuyang, CUI Ping, FANG Yulin, SU Baofeng. Diagnosis of grapevine leafroll disease severity infection via UAV remote sensing and deep learning[J]. Smart Agriculture, 2023, 5(3): 49-61. DOI：10.12133/j. smartag. SA202308013 (in Chinese with English abstract)

1　引　言

卷叶病会导致葡萄生长和品质下降，使葡萄酒的色泽、口感和风味变差[1]。及时准确地诊断卷叶病感染程度，对于防治卷叶病、提高葡萄果实品质和酿酒潜力具有重要意义[2]。

卷叶病的识别与诊断可及时发现和隔离感病植株，防止病毒扩散[3]。目前，卷叶病毒的诊断方法

收稿日期：2023-08-10

基金项目：宁夏回族自治区重点研发计划项目（2021BEF02017）

作者简介：刘易雪，研究方向为田间植物表型方法研究。E-mail：sunnyliu@nwafu.edu.cn

*通信作者：苏宝峰，博士，教授，研究方向为田间植物表型快速获取及应用。E-mail：bfs@nwsuaf.edu.cn

主要是反转录-聚合酶链反应（Reverse Transcription-Polymerase Chain Reaction, RT-PCR）[4]，然而，样本采集和处理的复杂性、检测结果的不稳定性、检测成本的高昂性、检测范围的局限性等问题难以在葡萄园中大面积推行这种卷叶病诊断方法[5]。因此，需要开发更准确、可以覆盖更大检测范围的卷叶病毒检测方法，以满足葡萄园管理的需求。

无人机遥感技术广泛应用于作物长势、营养状况、病虫害等监测中[6]。同时，便捷、廉价的无人机遥感数据可提供近实时的数据，为农田管理和决策提供依据，如灌溉调度、施肥量、病虫害防治等[7]。在作物病害监测方面，利用无人机遥感数据结合深度学习方法，可以实现小麦条锈病的检测以及不同程度的分类[8,9]，也实现了水稻白叶枯病的检测以及不同发病率的分类[10]。

样本不均衡是限制无人机遥感识别卷叶病的一个关键问题。葡萄园冠层的健康、轻度感染、重度感染样本的数量存在显著差异，导致深度学习模型在训练时倾向于样本数量较多的类别，而忽视样本数量少的类别，从而降低了卷叶病诊断的准确率和鲁棒性[11]。作物病害的发生和发展受到多种因素的影响，如气候条件、土壤环境、病原种类、传播途径等，导致了作物病害感染程度的分布不均匀[12]。同时，作物病害的防治措施也会影响作物病害感染程度的分布，如及时施用农药或移除感染植株，会减少轻度或重度感染的样本数量[13]。

本研究探讨了使用无人机遥感和深度学习技术进行冠层尺度的葡萄卷叶病感染程度诊断的方法，针对卷叶病不同感染程度类别不平衡的问题，通过使用GANformer生成多样化的图像数据，使用ArcFace损失函数平衡不同类别的权重，以及使用不同归一化方法和通道注意力机制的Swin Transformer模型进行集成，以期提高对葡萄卷叶病感染程度诊断的精度。

2 材料与方法

2.1 无人机遥感数据获取和处理

蛇龙珠是一种受卷叶病影响严重的葡萄品种。以蛇龙珠为研究对象，在宁夏回族自治区青铜峡市甘城子镇酿酒葡萄原产地保护区西鸽酒庄（105.9242°E，38.0714°N）获取田间数据。使用的无人机遥感平台为大疆精灵4 RTK版。在北京时间2020年10月9日12:00～14:00，晴朗天气下执行飞行任务。设置纵向重叠率85%，旁向重叠率为75%。无人机的飞行高度为60 m。在使用Pix4D mapper进行几何校准、地理配准之后，处理无人机获取的原始图像集生成的正射影像空间分辨率为每像素1.74 cm。

2.2 卷叶病严重程度评估标准

通过田间调查的方法完成酿酒葡萄卷叶病带病毒株的调查，评估了调查区域内的采样点相应的葡萄植株是否患病，将卷叶病严重程度评估标准分为3个类别，分别是健康、轻度和重度。健康类别表示植株没有任何卷叶病的症状；轻度类别指植株不超过二分之一的叶片有卷叶病症状；重度类别表示植株有二分之一以上的叶片有卷叶病症状。

使用RTK手持机记录采样点的地理信息，采用WGS-84坐标系作为定位坐标系。同时，采集葡萄行两端点的地理坐标以确定葡萄行的位置。如图1所示，在研究区域内获取1128个采样点信息，其中包括鉴定蛇龙珠冠层健康的采样点130个，卷叶病为轻度的采样点377个，重度的采样点621个。

2.3 GANformer模拟图像生成方法

为解决数据不平衡的问题，使用GANformer[14]模型来生成模拟图像，以增加数据集的多样性和数量。GANformer是一个基于Transformer的生成对抗网络，可以生成高分辨率和多样性的图像。GANformer模型由一个生成器G和一个判别器D组成，它们之间进行对抗学习，模型结构如图2所示。

2.3.1 基于GANformer的蛇龙珠冠层模拟图像生成

为使用GANformer模型来生成模拟图像，首先将每个类别的数据集按8：2划分为训练集和测试集。使用PyTorch 1.8.2框架实现GANformer模型，并在一台配备NVIDIA GeForce RTX 3090 GPU的计算机上进行训练和测试。输入一个512维度的随机

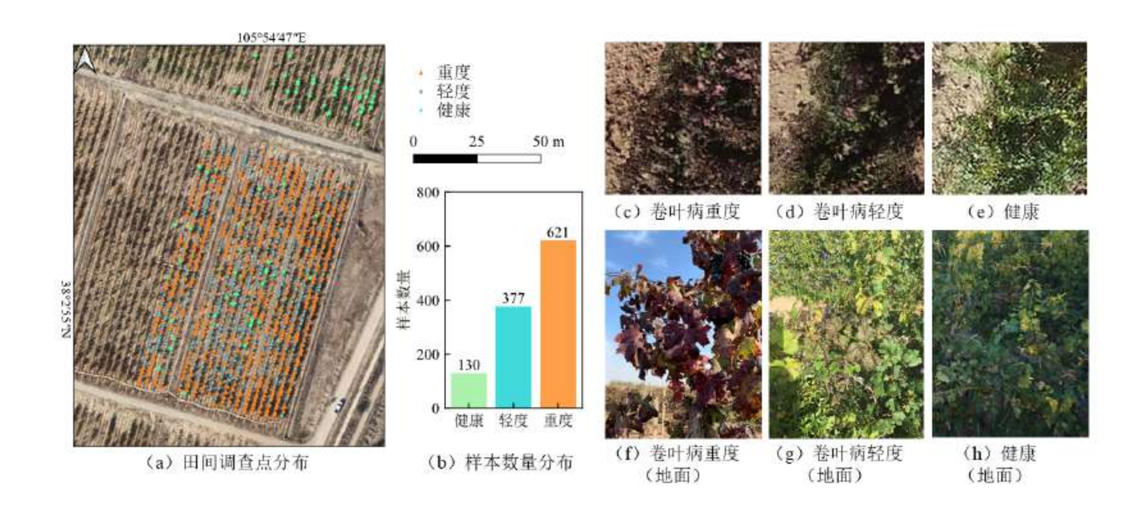


图1　宁夏西鸽酒庄葡萄园调查点分布及葡萄冠层患病不同等级的无人机遥感影像

Fig. 1 The location of field survey and grapevine canopy images of the grapevines with different levels infection in the vineyard of Xige Estate, Ningxia

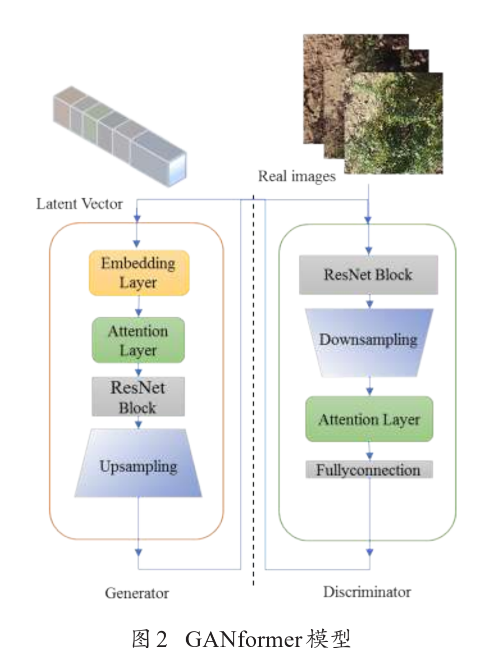


图2　GANformer模型

Fig. 2 GANformer model overview

噪声向量到生成器G，输出一个分辨率像素为256×256的RGB图像，同时判别器D输出一个标量值表示其真实性概率。使用Adam优化器进行优化，学习率为0.0002，动量参数分别为0.5和0.999。使用最小二乘损失（Least Squares Generative Adversarial Networks，LSGAN）作为损失函数[15]。模型使用渐进式训练策略，从88×88像素分辨率开始，每训练100个epoch就增加1倍的分辨率，直到达到256×256像素分辨率。每个分辨率的前50个epoch使用平滑过渡，以避免模式崩溃的现象。在训练过程中保存了每个分辨率下的生成器G的模型参数，并在测试过程中使用最高分辨率下的生成器G的模型参数。

2.3.2　模型评价方法

为评价GANformer模型生成模拟图像的效果，使用FID score来量化真实图像和模拟图像之间的差异[16]。加载一个预训练的Inception v3模型，并移除最后的分类层，将最后一个池化层的输出作为特征向量。对真实图像和生成图像分别计算它们的特征向量，并将它们分别拟合为一个多维高斯分布，得到它们的均值和协方差矩阵。最后，计算两个高斯分布之间的弗雷歇距离[17]，用公式（1）计算。

$$FID\ \text{score} = \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2 + \boldsymbol{Tr}\left(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 - 2\left(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2\right)^{\frac{1}{2}}\right) \tag{1}$$

其中，$\boldsymbol{\mu}_1$和$\boldsymbol{\mu}_2$分别是真实图像和生成图像的特征向量的均值；$\boldsymbol{\Sigma}_1$和$\boldsymbol{\Sigma}_2$分别是真实图像和生成图像的特征向量的协方差矩阵；$\boldsymbol{Tr}$表示矩阵的对角线元素之和。

2.4 基于Transformer的严重程度诊断模型

为评估蛇龙珠卷叶病的严重程度，本研究使用一种基于Transformer的深度学习分类网络模型，称为CA-Swin Transformer。该模型是在Swin Trans-

former [18] 的基础上进行改进，增加通道注意力机制和实例归一化层，同时使用ArcFace损失函数，以提高模型对蛇龙珠冠层图像的细粒度特征提取和下采样能力。改进后的CA-Swin Transformer结构如图3所示。

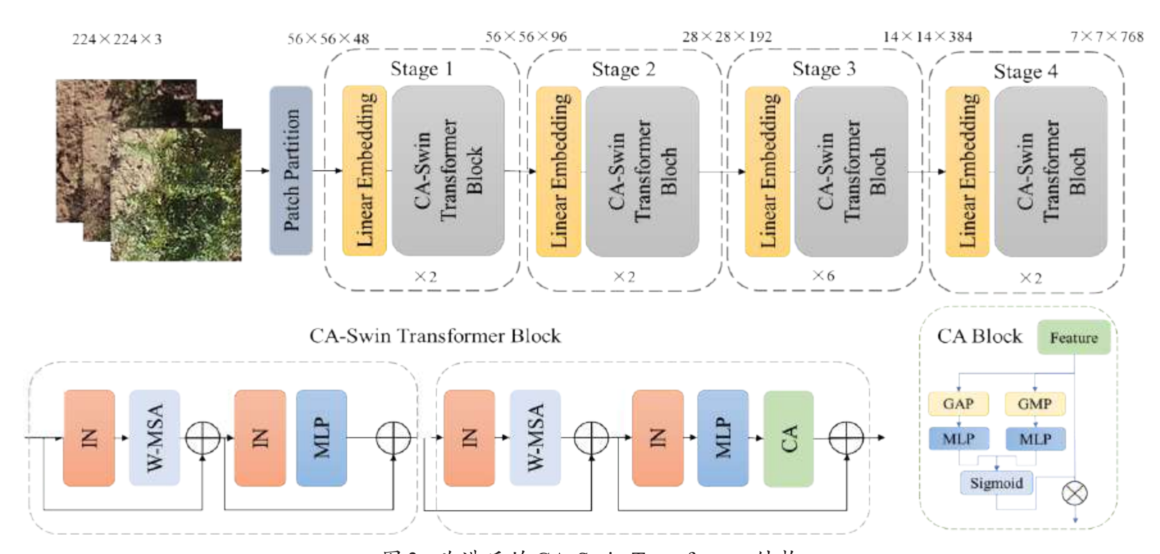


图3 改进后的CA-Swin Transformer结构

Fig. 3 The network architecture of the improved CA-Swin Transformer

2.4.1 CA-Swin Transformer模型

考虑到采集的蛇龙珠病害数据集的规模较小，本研究选用Swin Transformer tiny，将输入图像划分为4×4的小块，在每个小块上使用7×7大小的窗口来计算自注意力。共有4个阶段，每个阶段包含的Transformer layer的数量分别为2、2、6、2；每个变换器层使用的多头注意力的头数分别为3、6、12、24；多层感知器的隐藏层大小与输入层大小的比例设置为4。在蛇龙珠数据集上进行训练，输入图像的分辨率为224×224像素。

为提高图像分类模型的性能，本研究使用两种不同的归一化层：实例归一化层（Instance Normalization Layer，IN）和层归一化层（Layer Normalization Layer，LN）[19]。实例归一化层是对每个通道进行归一化操作，从而减少通道间的差异，并增强特征图的对比度；层归一化层是对整个样本进行归一化操作，从而捕捉不同通道之间的相关性，并增强特征图的表达能力。将这两种归一化层分别应用于Swin Transformer模型中，并在蛇龙珠数据集上进行实验。

此外，本研究还在每个Transformer层后增加一个通道注意力机制（Channel Attention，CA），用于提取不同通道之间的相关性，并增强特征图的表达能力。通道注意力机制由两个全连接层和一个激活函数组成，如公式（2）所示。

$$y = \sigma(W_2 \delta(W_1 x)) \tag{2}$$

其中，x为输入特征图；y为输出特征图；W_1和W_2为全连接层的权重矩阵；δ为GELU激活函数；σ是Sigmoid激活函数。将输出特征图与输入特征图相乘，得到经过通道注意力机制调整后的特征图。将这种改进后的Swin Transformer模型称为CA-Swin Transformer。

2.4.2 实施细节

为准确测试模型的泛化能力，使用留出法首先划分出测试集。测试集中健康、轻度卷叶病、重度卷叶病的样本数量分别为48、148、141。为保证训练集和验证集的数据分布一致性，按照8∶2的比例，分别对冠层健康、轻度、重度每个类别随机划分了原始数据集为训练集、验证集。从图1中可以发现，原始数据集存在明显的类别不平衡问题，即重度卷叶病样本的数量远多于健康和轻度卷叶病样本的数量。这会导致模型在训练过程中对重度卷叶病样本过拟合，而对其他类别样本欠拟合，从而降低模型在细粒度分类任务上的泛化能力。为缓解类别不平衡问题，使用2.3小节中GANsFormer模型来生成模拟图像，并将其与原始图像混合，从而增

强训练集的数据量和多样性。使用GANsFormer分别为健康、轻度卷叶病及重度卷叶病类别生成920、773和522张模拟图像，并将其与原始图像合并，使得每个类别的训练样本数量都达到1002张。

使用PyTorch框架实现CA-Swin Transformer模型，并在NVIDIA GeForce RTX 3090 GPU上进行训练和测试。使用Adam优化器来优化模型参数，并设置初始学习率为0.001，权重衰减为0.0001。使用余弦退火策略来调整学习率，并设置最小学习率为5×10^{-6}并设置批大小为32，训练轮数为100轮。使用早停法来防止模型过拟合，并设置验证间隔为10，容忍度为5。

分别使用交叉熵损失函数（Cross Entropy，CE）、Focal损失函数[20]和ArcFace损失函数[21]来计算模型的损失。交叉熵损失函数的定义为公式（3）。

$$L_{\mathrm{CE}} = -\frac{1}{N}\sum_{i=1}^{N} y_i \log \hat{y}_i \tag{3}$$

其中，L_{CE}为交叉熵损失；N为样本数，个；y_i为第i个样本的真实标签；$\hat{y}_i$为第i个样本的预测概率。

ArcFace损失函数在交叉熵损失函数的基础上增加一个角度边界，从而增强了类内紧密性和类间差异性[21]。ArcFace损失函数的定义为公式（4）。

$$L_{\mathrm{ArcFace}} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s(\cos(\theta_{y_i}+m))}}{e^{s(\cos(\theta_{y_i}+m))}+\sum_{j=1,j\neq y_i}^{n} e^{s\cos\theta_j}} \tag{4}$$

其中，L_{ArcFace}为ArcFace损失；s为特征向量的模长；θ_j为第i个样本的特征向量和第j个类别的权重向量之间的角度；m为一个超参数，表示角度边界。

Focal损失函数通过给难分类的样本赋予更大的权重，从而降低了易分类的样本对损失函数的贡献[20]。Focal损失函数的定义为公式（5）。

$$L_{\mathrm{Focal}} = -\frac{1}{N}\sum_{i=1}^{N}(1-\hat{y}_i)^{\gamma} y_i \log \hat{y}_i \tag{5}$$

其中，L_{Focal}为Focal损失；γ为一个超参数，表示难易分类样本的调节因子。

2.4.3　模型评价方法

为评价CA-Swin Transformer模型在蛇龙珠卷叶病严重程度评估任务上的性能，使用准确率（Accuracy）、召回率（Recall）、精确率（Precision）以及F_1分数（F_1-Score）4种评价指标。准确率表示模型预测正确的样本数占总样本数的比例；召回率表示模型预测正确的正样本数占实际正样本数的比例；精确率表示模型预测正确的正样本数占预测为正样本数的比例；F_1-Score表示准确率和召回率的调和平均值，用于综合评价模型的性能。在混淆矩阵中，当预测值和标签真值均为真时，即为预测为真的正样本（True Positives，TP）；当标签真值为真，预测值为假时，即为预测为假的负样本（False Negatives，FN）；当标签真值为假，预测值为真时，即为预测为假的正样本（False Positives，FP）；当标签真值为假，预测值为假时，即为预测为真的负样本（True Negatives，TN）。本研究分别计算了每个类别的准确率（公式（6））、精确率（公式（7））、召回率（公式（8））和F_1分数（公式（9）），以及所有类别的平均值。

$$Accuracy = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+FN}\times 100\% \tag{6}$$

$$Precision = \frac{\sum_{i=1}^{L}\frac{\mathrm{TP}_i}{\mathrm{TP}_i+\mathrm{FP}_I}\times w_i}{|L|} \tag{7}$$

$$Recall = \frac{\sum_{i=1}^{L}\frac{\mathrm{TP}_i}{\mathrm{TP}_i+\mathrm{FN}_I}\times w_i}{|L|} \tag{8}$$

$$F_1\text{-Score} = 2\times\frac{Precision_{\mathrm{weighted}}\times Recall_{\mathrm{weighted}}}{Preciosn_{\mathrm{weighted}}+Recall_{\mathrm{weighted}}} \tag{9}$$

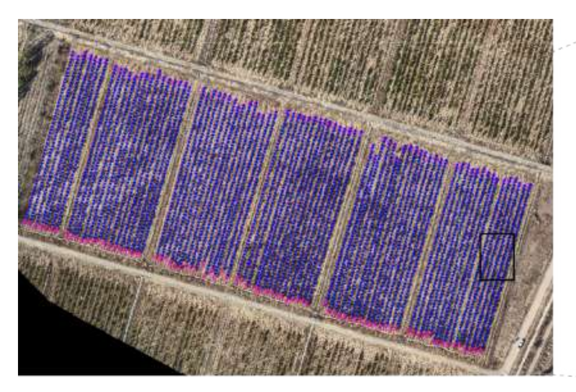
其中，w为类别数目权重；L为类别数目。

使用混淆矩阵来展示模型在每个类别上的预测结果，以及真实标签和预测标签之间的一致性。混淆矩阵的行表示真实标签；列表示预测标签；对角线上的元素表示预测正确的样本数量；非对角线上的元素表示预测错误的样本数量。

2.5 葡萄园卷叶病严重程度分布制图方法

为展示葡萄园卷叶病严重程度的分布情况，采用滑动窗口的方法，从正射影像中提取了蛇龙珠冠层的图像和标签。具体实施过程如图4所示。

根据葡萄行的两端位置确定滑动窗口的轨道，两端点确定了葡萄行所在直线方程。如图4所示，红色的线段表示葡萄行的轨道；黑色的点表示葡萄行的两端位置。根据葡萄株间距（1 m）得到了间



(a)在田块尺度使用滑动窗口轨道



(b)滑动窗口冠层分布细节

图4 在田块尺度使用滑动窗口遍历蛇龙珠冠层

Fig. 4 Grapevine canopy detection with slide windows at the field scale

隔的像素个数。蓝色的点表示每株葡萄的中心位置。根据株距确定滑动窗口大小为115×115像素确定滑动窗口的位置。记录滑动窗口的中心点横坐标x，中心点纵坐标y，窗口长度及宽度。中心点坐标为正射影像上的图像坐标。如图4所示，绿色的矩形表示滑动窗口的范围。

从正射影像中截取每个窗口内的图像，根据田间调查数据给每个窗口内的图像赋予标签。田间调查时对调查区域内的采样点相应的葡萄植株是否患病进行评估，并记录相应的位置信息。

在应用时，对由无人机图像生成的正射影像进行滑动窗口操作。设I为输入的RGB图像，其大小为$M \times H \times 3$，其中M和H分别为图像的高度和宽度；3为图像的通道数。首先将I转换为YCbCr颜色空间，得到I，其大小也为$M \times H \times 3$，然后得到以下3个阈值（公式（10）～（12））：

$$T_1 = (0.000, 0.504, 0.000, 0.504, 0.504, 1.000) \quad (10)$$

$$T_2 = (0.173, 0.647, 0.157, 0.635, 0.055, 0.478) \quad (11)$$

$$T_3 = (0.122, 0.162, 0.247, 0.473, 0.290, 0.635) \quad (12)$$

其中，T_1为用于分割土壤区域的阈值；T_2为用于分割冠层区域的阈值；T_3为用于分割卷叶病区域的阈值。每个阈值都包含3个通道的最小值和最大值。然后，使用公式（13）来计算每个像素是否属于土壤区域。

$$\boldsymbol{BW}_{\text{soil}}(i,j) = \begin{cases} 1 & \text{if} \begin{cases} I(i,j,1) \in [T_1(1), T_1(2)] \\ I(i,j,2) \in [T_1(3), T_1(4)] \\ I(i,j,3) \in [T_1(5), T_1(6)] \end{cases} \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

其中，$\boldsymbol{BW}_{\text{soil}}$为一个二值矩阵，其大小为$M \times H$，表示每个像素是否属于土壤区域，1表示“是”，0表示“否”；(i, j)表示像素的坐标；$(i, j, 1)$表示像素在Y通道上的值；$(i, j, 2)$表示像素在Cb通道上的值；$(i, j, 3)$表示像素在Cr通道上的值。

最后，使用公式（14）来计算土壤区域所占的比例。

$$P_{\text{soil}} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} \boldsymbol{BW}_{\text{soil}}(i,j)}{MH} \quad (14)$$

其中，P_{soil}为一个介于0和1之间的数值，表示土壤区域所占的比例。如果P_{soil}大于一个预设的阈值τ（$\tau = 0.3$），则认为窗口内没有冠层或冠层覆盖度很低；如果P_{soil}小于或等于τ，则认为窗口内有冠层。将判定有冠层的区域送入深度学习网络中进行分类预测，得到每个窗口内图像的预测标签。然后，根据预测标签将滑动窗口中心点显示出不同颜色，代表诊断结果。诊断结束后将预测结果的Python对象转换为shapefile文件，并使用QGIS软件完成制图。

3 结果与讨论

3.1 基于GANformer增强卷叶病严重程度诊断精度

3.1.1 GANformer模拟图像生成效果

图5显示了GANformer模拟图像生成过程中FID score随训练步数变化的曲线。从图5中可以看

出，FID score 从初始值 518.81 下降到最终值 102.60，在 1392 步左右达到最低点（93.20），并在后续波动稳定。这说明 GANformer 可以在较少的训练步数下生成高质量和多样性的图像。

图 6 显示了 GANformer 模拟图像生成效果中真实图像和模拟图像在不同训练步数下的视觉对比。从图 6 中可以看出，GANformer 生成的图像在视觉上与真实图像非常接近，能够保留原始图像中的颜色、纹理、形状等细节，并且能够生成不同的卷叶病严重程度的图像。

3.1.2　模拟图像对分类精度的影响

本研究使用 8 个不同的模型来对原始数据集和增强数据集进行分类训练，分别是 GoogLeNet [22]、MobileNetV2 [23]、NasNet Mobile [24]、ResNet18 [25]、ResNet50 [25]、CVT [26]、T2TViT [27] 和 Swin Transformer [18]。这些模型包括了卷积神经网络、注意力机制、Transformer 等不同的网络结构。本研究使用了准确率来评估不同模型上的分类效果，图 7 显示了不同模型在原始数据集和增强数据集上的准确率。

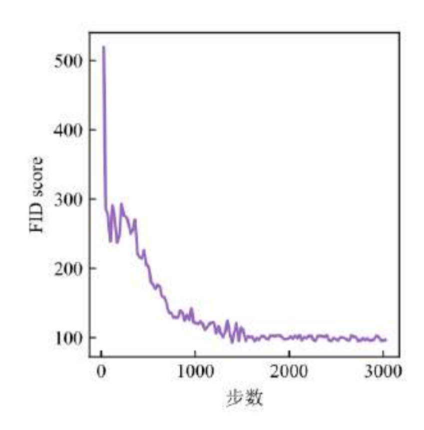


图 5　酿酒葡萄冠层无人机遥感模拟图像生成方法研究 GANformer 训练过程中 FID score 下降曲线

Fig. 5　FID score curve during GANformer training process for grapevine canopy simulated UAV image generation

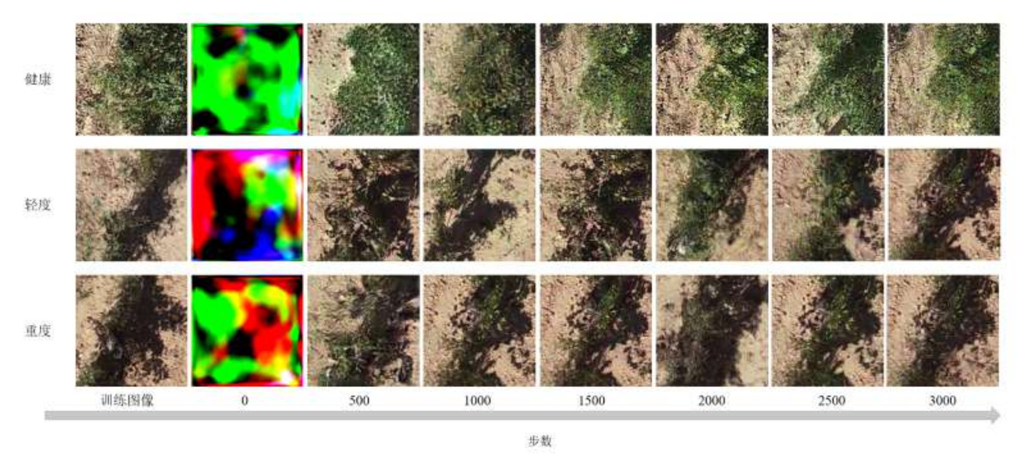


图 6　GANformer 模拟图像生成过程

Fig. 6 The image generation process of GANformer simulation

从图 7 中可以看出，增强数据集上的准确率普遍高于原始数据集上的准确率，说明 GANformer 生成的图像对于提升分类效果有正面的影响。一般来说，为了适应移动设备而设计的模型（如 GoogLeNet 和 MobileNetV2），在低分辨率图像上表现良好；而为了适应高分辨率的图像而设计的模型（如 NasNet Mobile 和 ResNet18），则可以在细节丰富和复杂的图像上表现优异；而使用 Transformer 结构的模型（如 CVT、T2TViT、Swin Transformer），则可以更好地利用 GANformer 生成的图像中的特征和信息，从而在复杂环境的图像上表现出色。

同时，从图 7 中可以看出，相比于使用卷积神经网络（Convolutional Neural Networks，CNN）结构的模型（如 GoogLeNet、ResNet 等），基于 Transformer 的深度学习模型在卷叶病严重程度分类的问题上更具备优势。这也验证了不同模型对于 GANformer 生成的图像有不同的适应性。最佳的模型是 Swin Transformer，在原始数据集的测试集上达到 80.11%。Swin Transformer 模型在增强数据集上的准确率为 83.97%，比原始数据集上提高 3.86%。

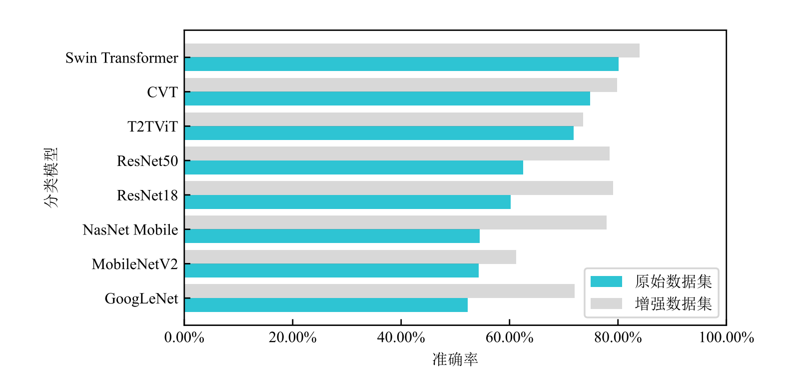


图7 使用GANformer增强酿酒葡萄冠层无人机遥感图像数据前后深度学习分类结果对比

Fig. 7 Comparison of deep learning classification results before and after using GANformer to enhance data for grapevine canopy simulated UAV image generation

为进一步分析GANformer模拟图像生成效果，本研究使用*t*分布随机邻域嵌入（*t*-Distributed Stochastic Neighbor Embedding，*t*-SNE）对学习到的特征进行了可视化。使用Swin Transformer提取了真实图像和模拟图像的特征，并使用*t*-SNE将它们降维到二维空间中[28]。图8显示了*t*-SNE对学习到的特征进行可视化的结果。

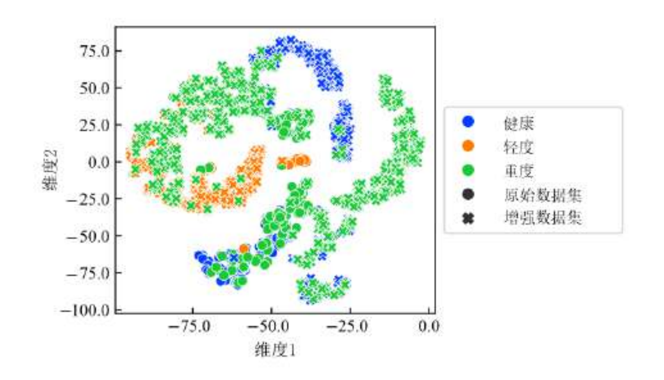


图8 酿酒葡萄冠层无人机遥感模拟图像生成方法研究基于*t*-SNE对模型学习到的冠层细粒度特征进行降维显示

Fig. 8 Dimensionality reduction display of canopy fine-grained- features learned from the model based on *t*-SNE for grapevine canopy simulated UAV image generation

从图8中可以看出，真实图像和模拟图像的特征在二维空间中有明显的聚类和分离效果，说明GANformer学习到的特征具有良好的结构和分布。真实图像和模拟图像在同一类别下的特征有较高的相似度，说明GANformer生成的图像能够保持原始图像中的类别信息。

3.2 CA-Swin Transformer诊断结果分析

表1显示了不同模型和方法在测试集上的分类精度，其中最高的精度用粗体标出。从表中可以看出，仅使用ArcFace代替交叉熵损失函数可以提升1.50%的分类精度，再使用IN作为归一化方法时，可再提升0.30%的准确率。都使用交叉熵作为损失函数，使用IN比使用LN提升1.20%的准确率。而使用ArcFace损失函数、实例归一化和通道注意力机制的CA-Swin Transformer模型在测试集上达到了最高的分类精度，为86.65%，比原始的Swin Transformer模型提高2.69%。这说明使用这些方法可以有效地提高模型对蛇龙珠卷叶病严重程度的识别能力。

图9显示了CA-Swin Transformer模型在测试集上的混淆矩阵，其中每个单元格表示预测类别和真实类别的匹配情况，颜色越深表示匹配程度越高。从图9中可以看出，尽管测试集的类别数量不平衡，使用ArcFace损失函数、实例归一化和通道注意力机制的CA-Swin Transformer模型在测试集上有最多的深色单元格，说明它能够更准确地将图像分为健康、轻度卷叶病和重度卷叶病3个类别。ArcFace损失函数是一种基于角度边界的损失函数，它可以增强特征空间中类别之间的区分度，提高分类效果。相比于传统的交叉熵损失函数，ArcFace损失函数可以更好地处理数据不平衡和类内方差大等问题，这一优势也体现在本研究的实验结果中。

为更详细地分析每个卷叶病感染类别的分类性能，表2展示了不同的模型和方法在测试集上不同

表1　不同改进方法在测试集上的分类效果

Table 1　Classification results of different improved methods on the test set

模型	损失函数	归一化方法	Precision	Recall	Accuracy/%	F_1-Score
Swin Transformer	交叉熵	LN	0.8415	0.8398	83.96	0.8401
Swin Transformer	Focal	LN	0.8445	0.8427	84.27	0.8432
Swin Transformer	ArcFace	LN	0.8598	0.8546	85.46	0.8548
Swin Transformer	交叉熵	IN	0.8623	0.8516	85.16	0.8509
Swin Transformer	ArcFace	IN	0.8621	0.8576	85.76	0.8573
CA-Swin Transformer	**ArcFace**	**IN**	**0.8692**	**0.8665**	**86.65**	**0.8669**

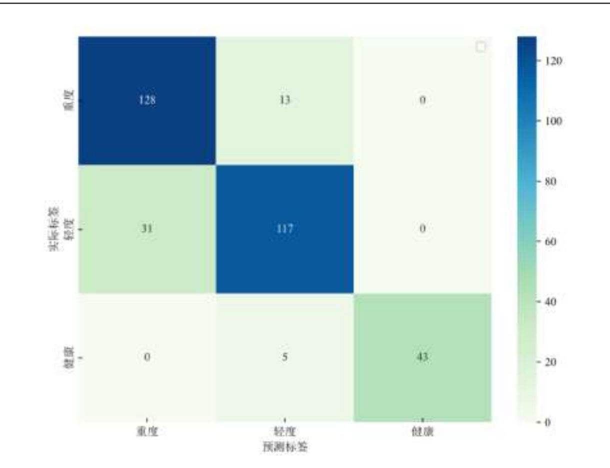


图9　卷叶病严重程度诊断的CA-Swin Transformer测试混淆矩阵

Fig. 9　CA-Swin Transformer confusion matrix on test for grapevine leafroll disease severity infection diagnosis

类别的分类结果。

表2　CA-Swin Transformer诊断结果测试集中不同卷叶病感染类别的分类效果

Table 2　CA-Swin Transforme classification results of different grapevine leafroll disease infection categories of the test set

等级	Precision	Recall	F_1-Score
健康	1.0000	0.9375	0.9677
轻度	0.8721	0.8226	0.8467
重度	0.8238	0.8851	0.8534

图10显示了基于Grad-CAM的不同模型和方法对应3个卷叶病感染类别的类激活图，其中每个子图表示模型对一个测试图像的分类结果和类激活区域，颜色越红表示区域越重要[29]。

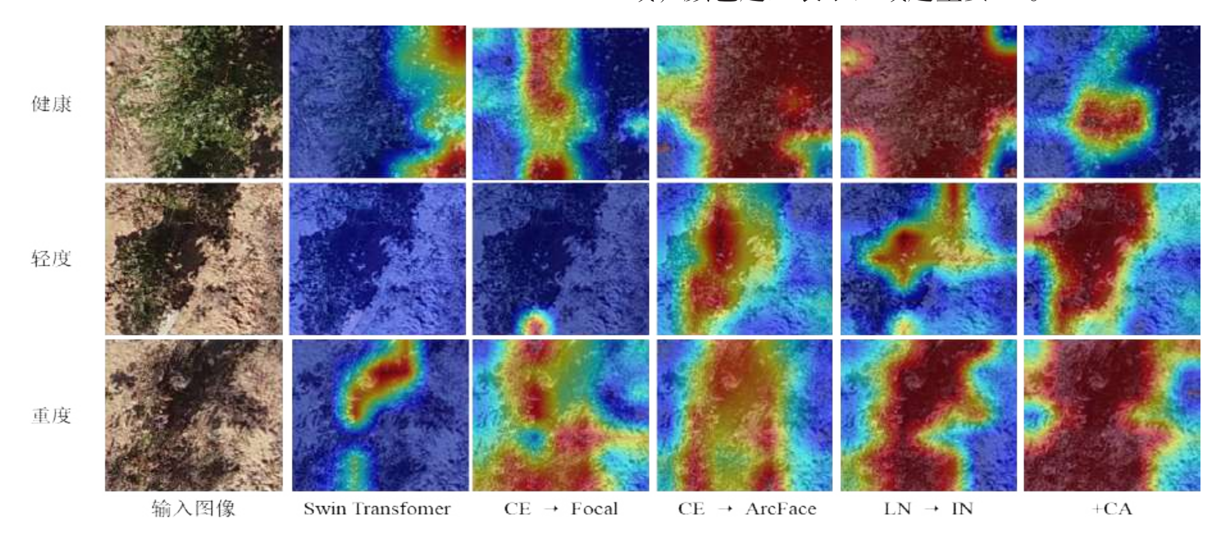


图10　基于Grad-CAM显示改进模型过程中的类激活图

Fig. 10　Class activation diagram based on Grad-CAM display model improvement

从图10中可以看出，使用ArcFace损失函数、Instance归一化和通道注意力机制的CA-Swin Transformer模型在类激活图上有最多的红色区域，说明它能够更好地捕捉图像中的细节和语义信息，从而提高分类效果。相反，原始的Swin Transformer模型在类激活图上有较少的红色区域，说明它有较多

的分类错误。相比于没有使用通道注意力机制的模型，使用通道注意力机制的模型可以更好地捕捉图像中的细节和语义信息，提高模型的识别能力。但是，通道注意力机制会增加模型的计算量和参数量，降低模型的效率。

针对分类类别不平衡的问题，本研究尤其关注模型在数量较少的类别上学习到的特征。从图10中可以看出，CA-Swin Transformer模型能够准确地将轻度、重度感染卷叶病的蛇龙珠冠层区分出来，并且在类激活图上显示出整个冠层区域都是重要的。相比于其他模型，在轻度和重度卷叶病类别上也有较高的分类性能，F_1分数分别达到0.8467和0.8534。从表2可以看出，CA-Swin Transformer模型F_1分数在健康类别上达到0.9677，这说明该模型能够将健康植株与感染卷叶病的植株区分开来。这一结果与图10中的类激活图是一致的，因为CA-Swin Transformer模型在健康类别上只关注了冠层区域，而忽略了其他无关的区域，从而提高了分类效果。相比于其他模型，初始的Swin Transformer在健康类别图像上显示阴影处是最重要的，而使用Focal损失函数则显示冠层的边缘是最重要的，使用ArcFace损失函数、实例归一化方法帮助了在轻度样本上的学习，然而在健康样本上却显示冠层和阴影都是很重要的，没有区分出最重要的特征。本研究的实验结果证明，使用通道注意力机制可以帮助基于Transformer的分类模型在数量较少的困难样本上学习到最重要的特征。

3.3 葡萄园卷叶病严重程度分布

图11显示了使用本研究所提方法制作的葡萄园蛇龙珠感染严重程度分布图。使用滑动窗口的方法来提取葡萄园冠层的图像和标签，将提取的图像送入训练好的CA-Swin Transformer网络中进行分类预测，得到每个窗口内图像的预测标签。根据预测标签将滑动窗口中心点显示出不同颜色，代表诊断结果莓红色表示重度卷叶病，湖蓝色表示轻度卷叶病，黄色表示健康。图11使用1：1250的比例尺来显示葡萄园的空间分布情况，为葡萄园卷叶病的防治提供了空间参考。

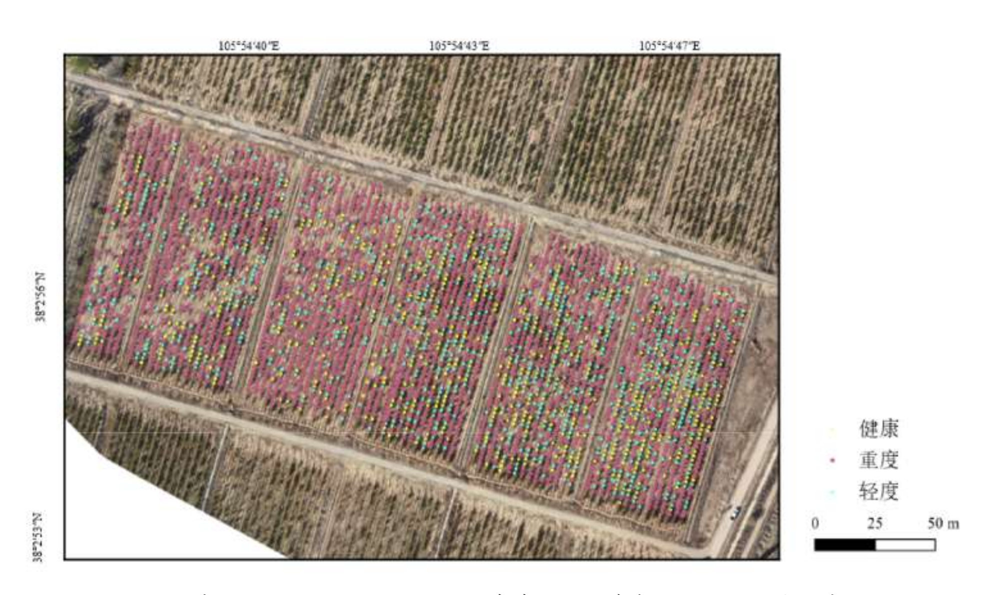


图11 基于CA-Swin Transformer的葡萄园蛇龙珠卷叶病严重程度分布

Fig. 11 Grapevine canopy detection with slide windows at the field scale based on CA-Swin Transformer

本研究通过制作葡萄园卷叶病严重程度分布图，发现：(1) 卷叶病主要分布在葡萄园中部和西部区域，东部区域蛇龙珠重度感染的数量少于西部区域。(2) 卷叶病严重程度与葡萄行之间存在一定的相关性，蛇龙珠感染卷叶病重度数量较多的区域越容易出现缺株、弱株情况。(3) 卷叶病严重程度与葡萄株之间也存在一定的差异。一般来说，葡萄行内部或两端的葡萄株卷叶病严重程度较高，而葡萄行中间或边缘的葡萄株卷叶病严重程度较低。

4 结　论

本研究以蛇龙珠为例，探讨了使用无人机遥感技术和深度学习进行冠层尺度的卷叶病感染程度诊断的方法，主要结论如下。

（1）针对蛇龙珠卷叶病不同感染程度类别不平衡的问题，提出了一种基于GANformer的数据增强方法。该方法可以生成蛇龙珠冠层模拟图像，从而增加了数据集的多样性，缓解了类别不平衡的问题。实验结果证明了该增强方法将Swin Transformer在卷叶病不同感染程度的分类精度提升3.86%。

（2）针对无人机遥感图像分类任务的特点，提出一种基于CA-Swin Transformer的图像分类方法。该方法通过引入通道注意力机制来增强特征表达能力，提高了模型的识别性能。相比于Swin Transformer，CA-Swin Transformer模型将分类精度提高了2.69%。

（3）针对蛇龙珠卷叶病感染程度分布图的制作需求，提出了一种基于滑动窗口和CA-Swin Transformer模型的方法。该方法通过提取葡萄园冠层的图像和标签，送入训练好的CA-Swin Transformer网络中进行分类预测，得到每个窗口内图像的预测标签，并根据预测标签将滑动窗口中心点显示出不同颜色，代表诊断结果。使用本研究所提方法制作的蛇龙珠卷叶病严重程度分布图，可以清晰地反映出葡萄园内不同区域、不同行列、不同株间的卷叶病感染情况，为葡萄园提供了卷叶病防治和精准化管理的依据。

利益冲突声明：本研究不存在研究者以及与公开研究成果有关的利益冲突。

参考文献：

[1] 刘万好, 于素珍, 肖慧琳, 等. 蛇龙珠卷叶病毒病原鉴定和对果实品质的影响[J]. 中国酿造, 2020, 39(11): 187-191.
LIU W H, YU S Z, XIAO H L, et al. Pathogen identification of Cabernet Gernischet grapevine leaf-roll virus and its effect on fruit quality[J]. China brewing, 2020, 39(11): 187-191.

[2] HOBBS M B, VENGCO S M, BOLTON S L, et al. Meeting the challenge of viral disease management in the US wine grape industries of California and Washington: Demystifying decision making, fostering agricultural networks, and optimizing educational resources[J]. Australian journal of grape and wine research, 2023, 2023: 1-17.

[3] ALMEIDA R P P, DAANE K M, BELL V A, et al. Ecology and management of grapevine leafroll disease[J]. Frontiers in microbiology, 2013, 4: ID 94.

[4] NAIDU R, ROWHANI A, FUCHS M, et al. Grapevine leafroll: A complex viral disease affecting a high-value fruit crop[J]. Plant disease, 2014, 98(9): 1172-1185.

[5] GAO Z M, KHOT L R, NAIDU R A, et al. Early detection of grapevine leafroll disease in a red-berried wine grape cultivar using hyperspectral imaging[J]. Computers and electronics in agriculture, 2020, 179: ID 105807.

[6] 兰玉彬, 邓小玲, 曾国亮. 无人机农业遥感在农作物病虫草害诊断应用研究进展[J]. 智慧农业, 2019, 1(2): 1-19.
LAN Y B, DENG X L, ZENG G L. Advances in diagnosis of crop diseases, pests and weeds by UAV remote sensing[J]. Smart agriculture, 2019, 1(2): 1-19.

[7] MAES W H, STEPPE K. Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture[J]. Trends in plant science, 2019, 24(2): 152-164.

[8] SCHIRRMANN M, LANDWEHR N, GIEBEL A, et al. Early detection of stripe rust in winter wheat using deep residual neural networks[J]. Frontiers in plant science, 2021, 12: ID 469689.

[9] RUAN C, DONG Y Y, HUANG W J, et al. Integrating remote sensing and meteorological data to predict wheat stripe rust[J]. Remote sensing, 2022, 14(5): ID 1221.

[10] XIAO D Q, PAN Y Q, FENG J Z, et al. Remote sensing detection algorithm for apple fire blight based on UAV multispectral image[J]. Computers and electronics in agriculture, 2022, 199: ID 107137.

[11] LÓPEZ V, FERNÁNDEZ A, GARCÍA S, et al. An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics[J]. Information sciences, 2013, 250: 113-141.

[12] SINGH A, JONES S, GANAPATHYSUBRAMANIAN B, et al. Challenges and opportunities in machine-augmented plant stress phenotyping[J]. Trends in plant science, 2021, 26(1): 53-69.

[13] FUCHS M. Grapevine viruses: A multitude of diverse species with simple but overall poorly adopted management solutions in the vineyard[J]. Journal of plant pathology, 2020, 102(3): 643-653.

[14] HUDSON D A, ZITNICK L. Generative adversarial Ttransformers[EB/OL]. arXiv: 2103.01209, 2021.

[15] MAO X D, LI Q, XIE H R, et al. Least squares generative adversarial networks[C]// 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway, New Jersey, USA: IEEE, 2017: 2813-2821.

[16] CHONG M J, FORSYTH D. Effectively unbiased FID and inception score and where to find them[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2020: 6069-6078.

[17] CHIZAT L, ROUSSILLON P, LÉGER F, et al. Faster was-

serstein distance estimation with the sinkhorn divergence[C]// Proceedings of the 34th International Conference on Neural Information Processing Systems. New York, USA: ACM, 2020: 2257-2269.
[18] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]// 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, New Jersey, USA: IEEE, 2022: 9992-10002.
[19] ULYANOV D, VEDALDI A, LEMPITSKY V. Instance normalization: The missing ingredient for fast stylization[EB/OL]. arXiv: 1607.08022, 2016.
[20] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]// 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway, New Jersey, USA: IEEE, 2017: 2999-3007.
[21] DENG J K, GUO J, XUE N N, et al. ArcFace: Additive angular margin loss for deep face recognition[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2020: 4685-4694.
[22] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2015: 1-9.
[23] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: Inverted residuals and linear bottlenecks[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, New Jersey, USA: IEEE, 2018: 4510-4520.
[24] ZOPH B, VASUDEVAN V, SHLENS J, et al. Learning transferable architectures for scalable image recognition[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, New Jersey, USA: IEEE, 2018: 8697-8710.
[25] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2016: 770-778.
[26] WU H P, XIAO B, CODELLA N, et al. CvT: Introducing convolutions to vision transformers[C]// 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, New Jersey, USA: IEEE, 2022: 22-31.
[27] YUAN L, CHEN Y P, WANG T, et al. Tokens-to-token ViT: Training vision transformers from scratch on ImageNet[C]// 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, New Jersey, USA: IEEE, 2022: 538-547.
[28] VAN DER MAATEN L, HINTON G. Visualizing data using t-SNE[J]. Journal of machine learning research, 2008. 9(11): 2579-2605
[29] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization[J]. International journal of computer vision, 2020, 128(2): 336-359.

Diagnosis of Grapevine Leafroll Disease Severity Infection via UAV Remote Sensing and Deep Learning

LIU Yixue[1,2,3], SONG Yuyang[4], CUI Ping[5], FANG Yulin[4], SU Baofeng[1,2,3*]

(1. *College of Mechanical and Electronic Engineering*, *Northwest A&F University*, *Yangling* 712100, *China*; 2. *Key Laboratory of Agricultural Internet of Things*, *Ministry of Agriculture and Rural Affairs*, *Yangling* 712100, *China*; 3. *Shaanxi Key Laboratory of Agriculture Information Perception and Intelligent Service*, *Yangling* 712100, *China*; 4. *College of Enology*, *Northwest A&F University*, *Yangling* 712100, *China*; 5. *Ningxia Helan Mountain East Foothill Wine Industry Park Management Committee*, *Yinchuan* 750002, *China*)

Abstract:

[Objective] Wine grapes are severely affected by leafroll disease, which affects their growth, and reduces the quality of the color, taste, and flavor of wine. Timely and accurate diagnosis of leafroll disease severity is crucial for preventing and controlling the disease, improving the wine grape fruit quality and wine-making potential. Unmanned aerial vehicle (UAV) remote sensing technology provides high-resolution images of wine grape vineyards, which can capture the features of grapevine canopies with different levels of leafroll disease severity. Deep learning networks extract complex and high-level features from UAV remote sensing images and per-

form fine-grained classification of leafroll disease infection severity. However, the diagnosis of leafroll disease severity is challenging due to the imbalanced data distribution of different infection levels and categories in UAV remote sensing images.

[Method] A novel method for diagnosing leafroll disease severity was developed at a canopy scale using UAV remote sensing technology and deep learning. The main challenge of this task was the imbalanced data distribution of different infection levels and categories in UAV remote sensing images. To address this challenge, a method that combined deep learning fine-grained classification and generative adversarial networks (GANs) was proposed. In the first stage, the GANformer, a Transformer-based GAN model was used, to generate diverse and realistic virtual canopy images of grapevines with different levels of leafroll disease severity. To further analyze the image generation effect of GANformer. The t-distributed stochastic neighbor embedding (t-SNE) to visualize the learned features of real and simulated images. In the second stage, the CA-Swin Transformer, an improved image classification model based on the Swin Transformer and channel attention mechanism was used, to classify the patch images into different classes of leafroll disease infection severity. CA-Swin Transformer could also use a self-attention mechanism to capture the long-range dependencies of image patches and enhance the feature representation of the Swin Transformer model by adding a channel attention mechanism after each Transformer layer. The channel attention (CA) mechanism consisted of two fully connected layers and an activation function, which could extract correlations between different channels and amplify the informative features. The ArcFace loss function and instance normalization layer was also used to enhance the fine-grained feature extraction and downsampling ability for grapevine canopy images. The UAV images of wine grape vineyards were collected and processed into orthomosaic images. They labeled into three categories: healthy, moderate infection, and severe infection using the in-field survey data. A sliding window method was used to extract patch images and labels from orthomosaic images for training and testing. The performance of the improved method was compared with the baseline model using different loss functions and normalization methods. The distribution of leafroll disease severity was mapped in vineyards using the trained CA-Swin Transformer model.

[Results and Discussions] The experimental results showed that the GANformer could generate high-quality virtual canopy images of grapevines with an FID score of 93.20. The images generated by GANformer were visually very similar to real images and could produce images with different levels of leafroll disease severity. The T-SNE visualization showed that the features of real and simulated images were well clustered and separated in two-dimensional space, indicating that GANformer learned meaningful and diverse features, which enriched the image dataset. Compared to CNN-based deep learning models, Transformer-based deep learning models had more advantages in diagnosing leafroll disease infection. Swin Transformer achieved an optimal accuracy of 83.97% on the enhanced dataset, which was higher than other models such as GoogLeNet, MobileNetV2, NasNet Mobile, ResNet18, ResNet50, CVT, and T2TViT. It was found that replacing the cross entropy loss function with the ArcFace loss function improved the classification accuracy by 1.50%, and applying instance normalization instead of layer normalization further improved the accuracy by 0.30%. Moreover, the proposed channel attention mechanism, named CA-Swin Transformer, enhanced the feature representation of the Swin Transformer model, achieved the highest classification accuracy on the test set, reaching 86.65%, which was 6.54% higher than using the Swin Transformer on the original test dataset. By creating a distribution map of leafroll disease severity in vineyards, it was found that there was a certain correlation between leafroll disease severity and grape rows. Areas with a larger number of severe leafroll diseases caused by Cabernet Sauvignon were more prone to have missing or weak plants.

[Conclusions] A novel method for diagnosing grapevine leafroll disease severity at a canopy scale using UAV remote sensing technology and deep learning was proposed. This method can generate diverse and realistic virtual canopy images of grapevines with different levels of leafroll disease severity using GANformer, and classify them into different classes using CA-Swin Transformer. This method can also map the distribution of leafroll disease severity in vineyards using a sliding window method, and provides a new approach for crop disease monitoring based on UAV remote sensing technology.

Key words: UAV remote sensing; deep learning; generate adversarial networks; Swin Transformer; leafroll disease of wine grape; data augmentation; attention mechanism

Foundation items: The Key R & D projects of Ningxia Hui Autonomous Region (2021BEF2017)